1.3 RECURSIVE BAYESIAN MODEL COMBINATION FOR STREAMFLOW FORECASTING

Newsha K. Ajami[1,*], Qingyun Duan[2], Hamid Moradkhni[1], Soroosh Sorooshian[1]
1. University of California at Irvine (UCI), Irvine, CA
2. Lawrence Livermore National Laboratory, Livermore, CA

1. Introduction

Operational and research hydrologic communities rely on a single model to predict streamflow. Yet, relying on a single model, which most likely does not adequately represents all the physical processes of the catchment, results in unreliable predictions. Recently there has been a movement towards producing consensus prediction by using probabilistic multi-model combination techniques such as Bayesian Model Averaging, BMA, (Hoeting et al., 1999, Fernandez, et al., 2001; Raftery et al., 2003). These Bayesian model combination approaches exploit the existing model structures and combine the model predictions based on their past performance.

Multi-model combination approaches such as BMA are fundamentally a special case of ensemble prediction techniques where ensemble members are the different model predictions. Various hydrologic model structures capture different aspects of the watershed response. It is commonly observed that hydrological models cannot represent different phases of the hydrograph equally well. Some models do well in simulating the peak flows, while others do well in capturing the low flows. This can be true in the case of a single model being applied to the river basin using different model parameters that are calibrated using criteria favoring high or low flows. One way of exploiting different model structures is to combine the outputs of various models using a judicious strategy that weigh the good performing models more favorably than the bad ones. This approach is designed to produce consensus predictions that are better than those by any single models. At the same time it provides a better description of the predictive uncertainty.

This study intends to introduce an ensemble forecast framework for hydrological forecast, called, Bayesian Recursive Model Combination (BRMC). In this framework, which is built over the foundation of Bayesian model combination techniques, the forecasts generated by various models are the ensemble members. These forecasts are being combined recursively based on their performance and uncertainty in reproducing the observation. The goal is to generate consensus forecast that is more skillful in all parts of the hydrograph than any individual model while reducing the uncertainty of the model predictions.

This paper is organized as follows. The first section we will present a brief description of the Bayesian Model Averaging technique. Subsequently a summary of the Bayesian framework and the experiment setup are explained. The final section presents preliminary results of this work and discusses the future direction for improvement.

2. Bayesian Model Averaging (BMA):

Let's consider a quantity y to be the forecasted variable and $M=[M_1, M_2, \ldots, M_K]$ the set of all considered models. $p_k(y|M_k, D)$ is the posterior distribution of y under model M_k, given a discrete data set, D. The posterior distribution of the BMA prediction is therefore given as:

$$p(y \mid M_1,...,M_k,D) = \sum_{k=1}^{K} p(M_k \mid D) \cdot p_k(y \mid M_k, D) \qquad (1)$$

where $p(M_k|D)$ is the posterior probability of model M_k. This term is also known as the likelihood of model M_k being the correct model. If we denote $w_k = p(M_k|D)$, we should obtain $\sum_{k=1}^{K} w_k = 1.$ Suppose that f_k is prediction made by model M_k. The posterior mean and variance of the BMA prediction for variable, y, are:

$$E[y \mid f_1..., f_K, D] = \sum_{k=1}^{K} w_k f_k \qquad (2)$$

$$Var[y \mid f_1..., f_K, D] = \sum_{k=1}^{K} w_k \left(f_k - \sum_{i=1}^{K} w_i f_i \right)^2 + \sum_{k=1}^{K} w_k \sigma_k^2 \qquad (3)$$

* Corresponding author address: Newsha K. Ajami, Univ. of California at Irvine, Dept. of Civil and Environmental Eng., E/4130 Engineering Gateway, Irvine, CA, 82697-2175

where σ^2_k is the variance of model M_k. In essence, the BMA prediction is the average of predictions weighted by the likelihood that an individual model is correct. There are several attractive properties to the BMA predictions. First the BMA prediction receives higher weights from better performing models as the likelihood of a model is essentially a measure of the agreement between the model predictions and the observations. Second, the BMA variance is a measure of the uncertainty of the BMA prediction. This term is a better measure of uncertainty because it accounts for both the between-model-variance and within-model-variance, as shown in the first and second terms of equation (3). Equations (2) and (3) can be solved iteratively by schemes such as the Expectation and Maximization scheme (Raftery et al, 2003) and the Markov Chain Monte Carlo (MCMC) methods (Gelman et al., 2004).

2.1Recursive Bayesian Model Combination (RBMC) for Streamflow Forecasting

Three conceptual rainfall-runoff models, Sacramento Soil Moisture Accounting model, SAC-SMA, (Burnash et al. 1973), Hydrologic MODel, HYMOD, (Boyle, 2000) and Simple Water Balance model, SWB, (Schaake et al., 1996) are used in this study. These models introduce different levels of complexity and have different skill and reliability levels. Five watersheds are selected from the study basins used in international MOdel Parameter Estimation Experiment (MOPEX). Each model is calibrated to historical streamflow observation available for each basin by applying Shuffled Complex Evolution (SCE; Duan et al. 1992) based on five different criteria (objective function) which emphasize on different parts of the hydrograph:

- Fitting the streamflow with higher stress on high flows by employing
    1. Sum of quadratic errors,
    2. Sum of square errors.
- Fitting both low and high flows
    3. Sum of absolute error
- Fitting the streamflow with higher stress on low flows by using
    4. Heteroscedastic Maximum Likelihood Estimator
    5. Sum of log of errors

Fifteen sets of flow simulations for each basin were generated using the calibrated model parameters (3 models x 5 parameter sets). Figures (1a-c) display the 15 model simulations. It can be observed that SAC predictions on average do well in simulating high flows, while HyMOD predictions do reasonably in representing the low flows. SWB predictions do extremely poor in predicting low flows. Given that the predictions have different skills in matching various sections of the hydrograph, we hope to obtain the consensus river flow prediction that extracts the strengths from each model prediction while avoiding the weaknesses. Two strategies were undertaken. First the combination weights (posterior probability of each model) were calculated over the entire time series (daily streamflow data from 1961 to 1965). In the second approach streamflow values were sorted in ascending order and then divided into ten different quantiles. Subsequently the weights are calculated separately for each quantile. The Expectation-Maximization (EM) algorithm is used to compute the weights. For brevity, EM is not presented here. Refer to Raftery et al. (2003) for details.

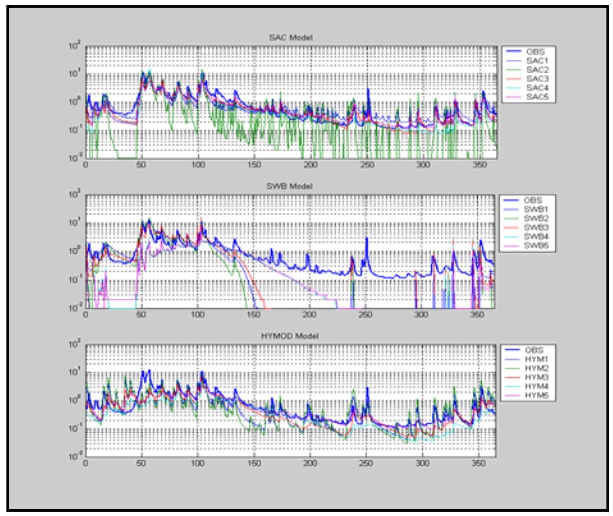


Figure (1) The streamflow simulations of the three hydrologic models using 5 different sets of calibrated parameters

3. Results and Discussion

The model combination results are first compared to any individual member model. Then, the results are compared among two different BRMC approaches; one with a single set of weights and the other one with multiple sets of weights. Finally, all of the results were compared to the Simple Model Averaging, SMA, where all the models are assigned equal weights. The questions we try to address from this numerical experiment are: (1) does BRMC predict streamflow better than any single model? (2) how does BRMC prediction compare to SMA prediction? (3) Do model weights reflect model performance?

Basins/ model	"01608 500"	"01643 000"	"01668 000"	"03054 500"	"03179 000"
SAC1	0.589	0.668	0.657	0.705	0.733
SAC2	0.727	0.551	0.775	0.770	0.784
SAC3	0.701	0.687	0.726	0.745	0.747
SAC4	0.593	0.58	0.627	0.574	0.605
SAC5	0.712	0.633	0.713	0.66	0.623
SWB1	0.513	0.508	0.576	0.656	0.738
SWB2	0.574	0.428	0.650	0.664	0.758
SWB3	0.488	0.503	0.663	0.642	0.74
SWB4	0.443	0.348	0.40	0.564	0.452
SWB5	0.474	0.333	0.381	0.367	0.474
HYM1	0.278	0.440	0.321	0.546	0.661
HYM2	0.549	0.222	0.516	0.587	0.728
HYM3	0.437	0.505	0.5	0.551	0.67
HYM4	0.437	0.337	0.375	0.472	0.548
HYM5	0.513	0.368	0.513	0.472	0.639
BRMC	0.534	0.525	0.661	0.657	0.767
BRMC-10q	0.742	0.737	0.787	0.789	0.846
SMA	0.702	0.677	0.715	0.714	0.770

Table (1): Efficiency of all member models, model combination strategies and simple model averaging to match the observation for all basins

Figure (2) presents the overall efficiency of the member models compared to BRMC prediction with 10 quantiles and SMA prediction. The figure indicates the superior performance of BRMC with 10 quantiles over any individual model. As one can see for all quantiles BRMC outperforms every single individual model.

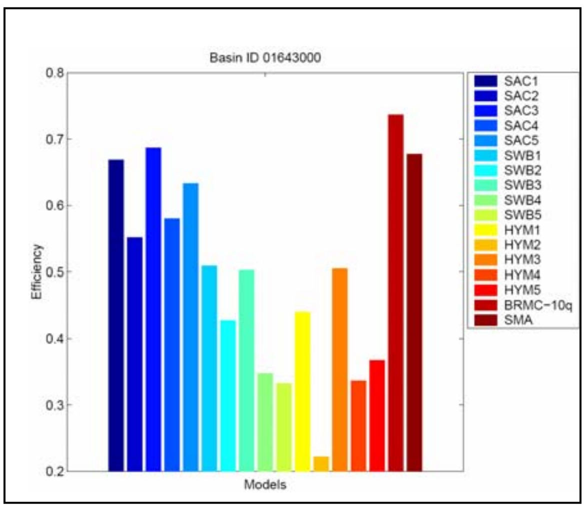


Figure (2): Overall Efficiency of the member models, BRMC with 10 quantiles and SMA for Basin number "01643000"

Table (1) presents the efficiency of each individual model, as well as both BRMC approaches and simple model averaging, with the observed stream flow. It is important to notice both BRMC approaches outperform any individual member model for all of the basins. The results demonstrate that dividing the time series into different flow levels considerably improves the results compared to only having one set of weights for the entire time series. Figure (3) proves this statement. One can clearly observe in the figure that BRMC with 10 quantile outperforms both simple BRMC as well as simple model averaging. This can be explained by considering the fact that just one set of weights for matching all flow levels makes it more arduous to account for bias (error) in all sections of the hydrograph. Whereas estimating separate sets of weights for each flow level enables us to match various parts of the hydrograph more accurately and decrease the uncertainty in the forecast. This can be confirmed by studying figures (4) and (5). Figure (4) shows the weights for each individual model over the entire time series, while figure (5) represents the weights for all models over each quantile.

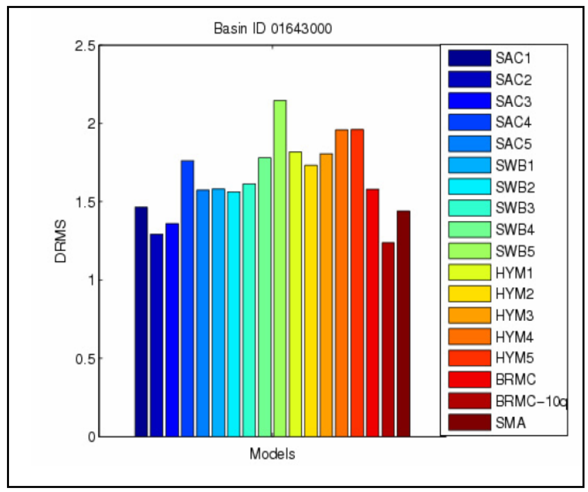
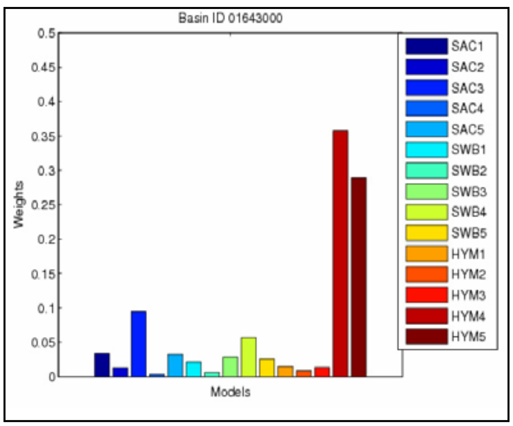


Figure (3) DRMS for all the individual models, Simple BRMC, BRMC with 10 quantiles and SMA

As one can see in figure (4), SAC-SMA and SWB models carry low weights compared to HYMOD models. This result indicates that these models do not have any significant role in matching the observed hydrograph. While looking at figure (5), one can see that the SAC 4 and SAC 5 models, which were calibrated using objective functions with more emphasis to match low flows (HMLE and LOG), carry significant weights within the low flow quantiles.

However, these models (SAC4 and SAC5) start carrying lower weights while SAC3 (with more emphasis on mid flows) holds higher weights as we move toward mid flows. Finally, SAC1 and SAC2 start picking up higher weights as we move toward higher flows. Hence, simple BRMC with one set of weights does not consider the skill of these aforementioned models through estimation of weights and assigns them all low weights. This shows the lack of reliability and skill for simple BRMS with just one set of weights and explains that for a more accurate streamflow forecast, dividing the hydrograph to different section is a necessary approach. Sectioning the hydrograph helped the weight estimating process assign multiple weights to the model simulations which were generated using parameter sets estimated applying different objective functions throughout the various quantiles (figure (5)). This confirms the consistency of the algorithm to capture the skill level of each member model. It also confirms that the weights carry a physical meaning and represents the model's performance within each quantile.



Figure(4): Weights of the individual model for simple BRMC (One set of weights over the entire time series)

In addition, all of the results were compared to the Simple Model Averaging method. As illustrated in Figures (2) and (3) as well the table (1), consensus forecast applying variable weights over 10 different quantiles consistently outperforms the SMA results. It is important to mention that using weights which vary from model to model makes more physical sense and decreases the uncertainty in the forecast. Assigning equal weights to all the models ignores the past performance as well as the innate uncertainty of the model.

These are the preliminary results of an ongoing study. The investigations on the number of models needed, the effect of the unknown probability distribution function of model errors, the application of Monte Carlo sampling techniques to deal with non-Gaussian distribution, the consideration of the input, model structure and the model parameters uncertainty, and finally how to quantify these uncertainties are part of the ongoing study. More comprehensive and complete results will be presented in a near future.

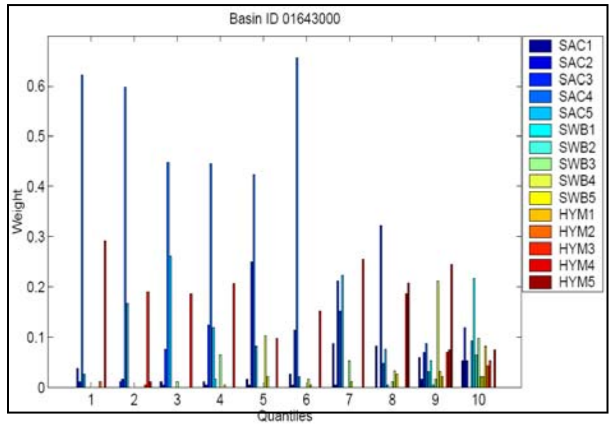


Figure (5): Weights of individual member model during different quantile for basin number "01643000"

Acknowledgments

This work was supported by NSF Sustainability of semi-Arid Hydrology and Riparian Areas (SAHRA) Science and Technology Center (NSF EAR-9876800) and HyDIS project (NASA grant NAG5-8503).

References

Boyle, D., Multicriteria calibration of hydrological models, Ph.D. thesis, Univ. of Arizona, Tucson, 2000.

Burnash, R.J., R.L. Ferral, and R.A. McGuire (1973). A Generalized Streamflow Simulation System Conceptual Modeling for Digital Computers, U.S. Department of Commerce National Weather Service and State of California Department of Water Resources.

Duan, Q., Sorooshian, S., and Gupta, V.K., 1992. Effective and Efficient Global Optimization for Conceptual Rainfall-Runoff Models. Water Resources Research, 28(4), 265-284.

Fernandez, C., Ley, E., and Steel, M., (2001). Benchmark Priors for Bayesian model averaging. Journal of Econometrics, 100, 381-427.

Hoeting, J.A., Madigan, D., Raftery, A.E., and Volinsky C.T., 1999. Bayesian Model Averaging: a tutorial. Statistical Science, 14(4), 382-417.

Raftery, A.E., and Zheng Y., 2003. Discussion : performance of Bayesian model averaging. Journal of American Statistical Association, 98(464), 931-938.

Schaake, J.C., V.I. Koren, Q.Y. Duan, K. Mitchell and F. Chen, 1996, Simple water balance model for estimating runoff at different spatial and temporal scales. J. Geophys. Res., 101(D3), 7461-7475